

Hi Koldo,

I wrote a variant that extracts a histogram of letter validity: low score=>bad letter.
The algorithm is mainly intended to detect near matches so it is not well suited to finding which letters are different (a basic compare loop should work better).

At least it reuses the "near match" information.

```
int correlation(const String& a, const String& b, int* aHistogram, int* bHistogram)
{
    int res=0;
    const char* A = a.Begin();
    const char* B = b.Begin();

    const int Al = a.GetLength();
    const int Bl = b.GetLength();
    const int deltaL = Al - Bl;

    int As, Bs, intersectLength;
    int subRes;

    int matchPatternMinLength = max(2, min(b.GetLength(), a.GetLength())/3);

    // init de l'histogramme
    for (int c = 0; c<Al; c++)
    {
        aHistogram[c] = 0;
    }

    for (int c = 0; c<Bl; c++)
    {
        bHistogram[c] = 0;
    }

    for (int offset = (matchPatternMinLength-Bl); offset <= (Al-matchPatternMinLength); offset++)
    {
        // determination des bornes
        if (offset < 0)
        {
            As = 0;
            Bs = -offset;
            if (offset < deltaL) intersectLength = offset + Bl;
            else                intersectLength = Al;
        }
    }
}
```

```

}
else
{
  As = offset;
  Bs = 0;
  intersectLength = Al - offset;
  if (offset <= deltaL) intersectLength = Bl;
  else          intersectLength = Al - offset;
}

subRes = 0;
for (int c = 0; c < intersectLength; c++)
{
  if( A[As+c] == B[Bs+c]) ++subRes;
}
if (subRes >= matchPatternMinLength)
{
  res += subRes;

  // creation de l'histogramme
  for (int c = 0; c < intersectLength; c++)
  {
    if( A[As+c] == B[Bs+c])
    {
      aHistogram[As+c] += subRes;
      bHistogram[Bs+c] += subRes;
    }
  }
}

// prise en compte du nbr de caracteres
if ( deltaL > 0 )
{
  res -= deltaL;
}
else
{
  res += deltaL;
}
return res;
}

```

Here are some test results:

COMPARE (123456789 , 123456789) corr= 9 ==> NEAR_MATCH
Histogram: - 9 - 9 - 9 - 9 - 9 - 9 - 9 - 9 - 9
Histogram: - 9 - 9 - 9 - 9 - 9 - 9 - 9 - 9 - 9

COMPARE (123456789 , 12345678) corr= 7 ==> NEAR_MATCH
Histogram: - 8 - 8 - 8 - 8 - 8 - 8 - 8 - 8 - 0
Histogram: - 8 - 8 - 8 - 8 - 8 - 8 - 8 - 8 - 8

COMPARE (123456789 , 1234567) corr= 5 ==> NEAR_MATCH
Histogram: - 7 - 7 - 7 - 7 - 7 - 7 - 7 - 0 - 0
Histogram: - 7 - 7 - 7 - 7 - 7 - 7 - 7 - 7 - 7

COMPARE (123456789 , 123456) corr= 3 ==> NEAR_MATCH
Histogram: - 6 - 6 - 6 - 6 - 6 - 6 - 0 - 0 - 0
Histogram: - 6 - 6 - 6 - 6 - 6 - 6 - 6 - 6 - 6

COMPARE (123456789 , 12345) corr= 1 ==> -----
Histogram: - 5 - 5 - 5 - 5 - 5 - 0 - 0 - 0 - 0
Histogram: - 5 - 5 - 5 - 5 - 5 - 5 - 5 - 5 - 5

COMPARE (123456789 , 1234) corr= -1 ==> -----
Histogram: - 4 - 4 - 4 - 4 - 0 - 0 - 0 - 0 - 0
Histogram: - 4 - 4 - 4 - 4 - 4 - 4 - 4 - 4 - 4

COMPARE (123456789 , 123) corr= -3 ==> -----
Histogram: - 3 - 3 - 3 - 0 - 0 - 0 - 0 - 0 - 0
Histogram: - 3 - 3 - 3 - 3 - 3 - 3 - 3 - 3 - 3

COMPARE (123456789 , 12) corr= -5 ==> -----
Histogram: - 2 - 2 - 0 - 0 - 0 - 0 - 0 - 0 - 0
Histogram: - 2 - 2 - 2 - 2 - 2 - 2 - 2 - 2 - 2

COMPARE (123456789 , 1) corr= -8 ==> -----
Histogram: - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0
Histogram: - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0

COMPARE (123456789 , 12345 89) corr= 7 ==> NEAR_MATCH
Histogram: - 7 - 7 - 7 - 7 - 7 - 0 - 0 - 7 - 7
Histogram: - 7 - 7 - 7 - 7 - 7 - 7 - 0 - 0 - 7 - 7

COMPARE (123456789 , 1234589) corr= 5 ==> NEAR_MATCH
Histogram: - 5 - 5 - 5 - 5 - 5 - 0 - 0 - 2 - 2
Histogram: - 5 - 5 - 5 - 5 - 5 - 5 - 2 - 2

COMPARE (123456789 , 12 56789) corr= 7 ==> NEAR_MATCH
Histogram: - 7 - 7 - 0 - 0 - 7 - 7 - 7 - 7 - 7
Histogram: - 7 - 7 - 0 - 0 - 7 - 7 - 7 - 7 - 7

COMPARE (123456789 , 1256789) corr= 5 ==> NEAR_MATCH
Histogram: - 2 - 2 - 0 - 0 - 5 - 5 - 5 - 5 - 5
Histogram: - 2 - 2 - 5 - 5 - 5 - 5 - 5

=====
COMPARE (123 , 123) corr= 3 ==> NEAR_MATCH
Histogram: - 3 - 3 - 3
Histogram: - 3 - 3 - 3

COMPARE (123 , 234) corr= 2 ==> NEAR_MATCH
Histogram: - 0 - 2 - 2
Histogram: - 2 - 2 - 0

COMPARE (123 , 345) corr= 0 ==> -----
Histogram: - 0 - 0 - 0
Histogram: - 0 - 0 - 0

COMPARE (123 , 3456) corr= -1 ==> -----
Histogram: - 0 - 0 - 0
Histogram: - 0 - 0 - 0 - 0

COMPARE (123 , 2345) corr= 1 ==> -----
Histogram: - 0 - 2 - 2
Histogram: - 2 - 2 - 0 - 0

COMPARE (123 , 01234) corr= 1 ==> -----
Histogram: - 3 - 3 - 3
Histogram: - 0 - 3 - 3 - 3 - 0

COMPARE (123 , 120023) corr= 1 ==> -----
Histogram: - 2 - 4 - 2
Histogram: - 2 - 2 - 0 - 0 - 2 - 2

COMPARE (SMITH , SMITH) corr= 5 ==> NEAR_MATCH
Histogram: - 5 - 5 - 5 - 5 - 5
Histogram: - 5 - 5 - 5 - 5 - 5

COMPARE (SMITH , SMITHE) corr= 4 ==> NEAR_MATCH
Histogram: - 5 - 5 - 5 - 5 - 5
Histogram: - 5 - 5 - 5 - 5 - 5 - 0

COMPARE (SMITH , SMISS) corr= 3 ==> NEAR_MATCH
Histogram: - 3 - 3 - 3 - 0 - 0
Histogram: - 3 - 3 - 3 - 0 - 0

COMPARE (SMITH , SMIF) corr= 2 ==> NEAR_MATCH
Histogram: - 3 - 3 - 3 - 0 - 0

Histogram: - 3 - 3 - 3 - 0

COMPARE (SMITH , SMICH) corr= 4 ==> NEAR_MATCH

Histogram: - 4 - 4 - 4 - 0 - 4

Histogram: - 4 - 4 - 4 - 0 - 4

COMPARE (SMITH , SWITH) corr= 4 ==> NEAR_MATCH

Histogram: - 4 - 0 - 4 - 4 - 4

Histogram: - 4 - 0 - 4 - 4 - 4

COMPARE (SMITH , SMATH) corr= 4 ==> NEAR_MATCH

Histogram: - 4 - 4 - 0 - 4 - 4

Histogram: - 4 - 4 - 0 - 4 - 4

COMPARE (SMITH , CMITE) corr= 3 ==> NEAR_MATCH

Histogram: - 0 - 3 - 3 - 3 - 0

Histogram: - 0 - 3 - 3 - 3 - 0

COMPARE (SMITH , CEMISS) corr= 1 ==> -----

Histogram: - 0 - 2 - 2 - 0 - 0

Histogram: - 0 - 0 - 2 - 2 - 0 - 0

COMPARE (REPETITION , REPETITION) corr= 16 ==> NEAR_MATCH

Histogram: - 10 - 13 - 10 - 13 - 13 - 13 - 13 - 13 - 10 - 10

Histogram: - 10 - 13 - 10 - 13 - 13 - 13 - 13 - 13 - 10 - 10

COMPARE (REPETITION , REPETITION) corr= 13 ==> NEAR_MATCH

Histogram: - 9 - 13 - 9 - 13 - 13 - 4 - 9 - 9 - 9 - 9

Histogram: - 9 - 9 - 9 - 13 - 9 - 4 - 13 - 13 - 9 - 9

COMPARE (REPETITION , REPETITION) corr= 11 ==> NEAR_MATCH

Histogram: - 8 - 11 - 8 - 11 - 8 - 3 - 0 - 8 - 8 - 8

Histogram: - 8 - 8 - 8 - 11 - 8 - 3 - 0 - 11 - 8 - 8

COMPARE (REPETITION , REPETITION) corr= 11 ==> NEAR_MATCH

Histogram: - 8 - 11 - 8 - 11 - 8 - 3 - 0 - 8 - 8 - 8

Histogram: - 8 - 8 - 8 - 11 - 8 - 3 - 0 - 11 - 8 - 8

COMPARE (REPETITION , REPETITION) corr= 7 ==> NEAR_MATCH

Histogram: - 5 - 5 - 5 - 5 - 5 - 0 - 0 - 3 - 3 - 3

Histogram: - 5 - 5 - 5 - 5 - 5 - 0 - 0 - 0 - 3 - 3 - 3

COMPARE (REPETITION , REPETITION) corr= 11 ==> NEAR_MATCH

Histogram: - 6 - 6 - 6 - 9 - 6 - 6 - 3 - 6 - 3 - 3

Histogram: - 6 - 9 - 6 - 6 - 9 - 9 - 0 - 0 - 3 - 3 - 3

COMPARE (REPETITION , RIPITITION) corr= 9 ==> NEAR_MATCH

Histogram: - 4 - 0 - 4 - 0 - 4 - 7 - 3 - 6 - 3 - 3

Histogram: - 4 - 0 - 4 - 3 - 7 - 7 - 0 - 0 - 3 - 3 - 3

COMPARE (REPETITION , RIPPITTISSION) corr= 4 ==> -----

Histogram: - 4 - 0 - 4 - 0 - 0 - 0 - 4 - 7 - 3 - 3

Histogram: - 4 - 0 - 4 - 0 - 0 - 0 - 4 - 4 - 0 - 0 - 3 - 3 - 3

COMPARE (REPETITION , RIKKITISSION) corr= 4 ==> -----

Histogram: - 0 - 0 - 0 - 0 - 0 - 3 - 3 - 6 - 3 - 3

Histogram: - 0 - 0 - 0 - 0 - 3 - 3 - 3 - 0 - 0 - 3 - 3 - 3